**FCND DISCUSSION PAPER NO. 43**

# HOW RELIABLE ARE GROUP INFORMANT RATINGS? A TEST OF FOOD SECURITY RATING IN HONDURAS

Gilles Bergeron, Saul Sutkover Morris, and Juan Manuel Medina Banegas

**Food Consumption and Nutrition Division**

**International Food Policy Research Institute**
**1200 Seventeenth Street, N.W.**
**Washington, D.C. 20036-3006 U.S.A.**
**(202) 862-5600**
**Fax: (202) 467-4439**

**April 1998**

## ABSTRACT

The reliability of food security rating, a variant of the more familiar wealth ranking procedure, was tested in a rural area of Western Honduras. Twenty workshop sessions were conducted in 13 different communities, with members of organized small farmers' groups attended by a large agricultural development project. Participants were all poor farmers with no more than 10 hectares of land. Participants, who generally knew each other well, were split into small sets and each set was asked to rate the food security status of all households in their organized group. Inter-rater agreement was assessed using the Kappa statistic, and related to other characteristics of the informants and workshop sessions, using multiple regression methods. Agreement was very poor (median value $\delta$ = 0.29), especially for the category "Intermittenly Food Insecure," and was associated with time elapsed since training of the session moderators. Women were 49 percent more likely than men to classify a given family as food insecure ($P < 0.001$). The authors put forward seven different hypotheses to explain the poor reliability of the rating method, which should be investigated in future research if the credibility of the method is to be reinforced.

## CONTENTS



## TABLES



## FIGURES

## ACKNOWLEDGMENTS

This research was made possible thanks to the generous contribution of the International Fund for Agricultural Development (IFAD). The comments of Alayne Adams, John Pender, Lawrence Haddad, Sara Scherr, and anonymous reviewers are gratefully acknowledged.

Gilles Bergeron and Saul Sutkover Morris
International Food Policy Research Institute

Juan Manuel Medina Banegas
Proyecto de Seguridad Alimentaria

# 1. INTRODUCTION

Group informant rating methods, such as wealth ranking, are increasingly popular among development workers, who use such tools to identify and target groups for specific activities and to understand local criteria of wealth and social mobility. It has been said that rating by local informants can help remove the biases of conventional survey methods because it brings intangible elements (such as status, and access to power or to networks of support) to the measurement of wealth and poverty, thus bridging the gap between outsider and local perceptions of poverty. The method is further credited with greatly reducing costs in identifying the poorest segments of a community. In addition, it is believed that the simplicity of the method makes it easy to teach and replicate by local people and development workers (Grandin 1983, 1994; Pretty et al. 1992; Gujit 1992a; Mukherjee 1992).

Various approaches have been used for conducting group informant ratings, but the most common one is the "index card" method.[1] Card ranking usually consists of writing the name of each household on a separate index card and having the informants sort them on some reference criteria (e.g., from poorest to wealthiest). Beyond this basic

[1]"Social maps" are also used, based on sketch maps drawn by villagers. Each household is located on this map and informants indicate its wealth status by putting a special marker on it. This approach is popular because it is quick, permits the spatial location of households, and facilitates the assignment of other characteristics (e.g., oxen ownership, water access, etc.) using distinct types of markers (Schaefer 1992; Mukherjee 1992). It does not explicitly rank units against one another, however, which may be why the card ranking method is usually preferred.

methodological principle, however, few conventions guide the use of this instrument. In spite of a number of well-documented case studies (e.g., Grandin 1983, 1994; IIED 1992), there is no conclusive evidence yet on, for example, exactly how informants should be selected and how many they should be, whether there are advantages to segregating them along social stratification criteria (sex, caste, etc.), and how much training, and of what kind, session moderators need to receive. As with other participatory appraisal approaches, it is possible that the rapid adoption of group informant rating methods has obscured the need to critically examine the conditions under which valid and reliable results can be generated.

The *validity* of the wealth ranking method was tested recently by Adams and colleagues with satisfactory results being reported (Adams et al. 1997). In another report, however, Chadwick and colleagues mention that the wealth ranking method was not found to be "a precise tool for establishing the socioeconomic profile of an area"; that social differences (gender and caste) affected ratings; and that the criteria for evaluating wealth tended to differ between groups (Chadwick et al. 1995). Similar findings of gender differences were reported by Wellbourn (1992). Besides, Scoones (1992) and Mearns and colleagues (1992) both find, in separate studies, that wealth rankings are susceptible to variations according to the criteria used, and thus conclude that the method may best be used as a complement to more conventional survey approaches.

All of the latter findings refer to the reliability of the method, rather than to its validity. Reliability concerns the extent to which a measuring procedure yields the same

results on repeated trials and is thus an assessment of the amount of random error associated with a particular procedure (Carmines and Zeller 1983). To the best of our knowledge, empirical tests of the *reliability* of the approach have not been proposed yet.

In this paper, we use data from exercises conducted among rural households in Honduras to assess the reliability of group informant ratings. In this case, informants were asked to classify households according to their ability to access food[2] (food security rating) rather than according to their wealth. Because the reliability assessments were replicated in 20 different settings, we were also able to examine three factors we believe may influence the reliability of the group informant rating method, namely (1) the selection of informants, (2) the effect of gender on informant rating behavior, and (3) the training of field personnel. Our findings lead us to hypothesize additional sources of unreliability in group informant ratings, establishing the basis for further testing of the method. We conclude that while the flexibility of the informant rating method is one of its merits (Chambers 1997), there may be simple principles that, if followed, could strengthen its reliability.

[2] Food security is defined as a situation where all people at all times have both physical and economic access to sufficient food to meet their dietary needs for a productive and healthy life (Frankenberger 1992). The concept of food security encompasses three distinct but related dimensions: food availability (the presence of food for all within a given geographical area); food access (capacity of households and individuals to obtain appropriate foods); and food utilization (biological capacity of individuals to effectively use the food consumed) (Riely et al. 1995). In this paper, we concentrate on the second of these dimensions, food access.

## 2. FOOD SECURITY RATING IN WESTERN HONDURAS: METHOD

Our study was designed to assess household food security (HFS) status in 13 communities in western Honduras (departments of Copán, Lempira, and Ocotepeque) using wealth ranking-like methods. The study took place in the context of a large agricultural development project in which village-level beneficiary groups were already identified. For that reason, only households targeted by the development project were included. Project beneficiaries were all members of organized small farmers' groups, and most participant households had a long history of involvement in local development projects, had known the other members of their beneficiary group for a long time, and shared kinship links with some of them. Since project beneficiaries obtained access to credit using group guarantees, there was a strong incentive for groups to exclude individuals whose creditworthiness (and thus, by implication, general living conditions) was unknown to other group members. In addition, a 10-hectare land ownership ceiling was enforced by the development project. For all these reasons, we believe it is safe to assume that the families observed represented a rather homogeneous group of “poor” households, and knew one another well.

The study took place over a period of six weeks. Every community was visited once for one day, during which time the participatory workshop was held. The international staff who designed the project were present during the first two weeks, after which time national staff were left to conduct the workshops on their own. All adults in the beneficiary households were invited to participate in the workshops. In 7 of the 13

villages, large numbers of women turned up to the workshop and were separated from the men[3]; in six other villages, no women came to the workshop. In the remaining community, a small number of women turned up and were integrated with the men. These 20 groups of informants were next divided into smaller sets of three to six persons. Informants could select their partners, the only recommendation being that each set must have at least one member that could read. Between 2 and 15 such small sets were formed, depending on the number of beneficiaries in the group (14 to 32 families), for a total of 123 sets across the 13 villages (see Table 2). This approach permitted us to reproduce over a number of replicates the single-group assessment usually employed in group ratings.

The workshop began with an explanation of the purpose of the exercise. It was made clear to informants that no particular benefits would accrue to them from participating in the exercise, and that it was essentially an appendage to the program that already brought them benefits.

Each small set of informants was asked to classify all the households belonging to their beneficiary group into one of three categories of household food security, which were predefined and carefully explained to the informants before the beginning of the session. The definition of the concept was first explained, then a locally suitable word was elicited to represent that concept. The categories were

[3]The findings cannot be directly compared in one community because inadvertently a different subset of beneficiaries were rated by men and women.

| Concept | Definitions |
|---|---|
| • food secure: | always have sufficient food on the table; |
| • intermittently insecure: | usually have sufficient food, but may have problems accessing or obtaining food in some years; |
| • food insecure: | have problems accessing or obtaining food every year. |

The content and construct validity of this three-way classification of food sufficiency has previously been established among relatively poor Spanish-speaking survey respondents in the United States (Briefel and Woteki 1992), and was also supported by preliminary discussions with project beneficiaries.

Each household had its name written on an index card, and every small set of informants was given a complete pack of cards representing all households in their beneficiary group. Most participants in the study were literate, and those that were not were paired with someone who was during the ranking exercise. Within each small set, informants were then invited to discuss the circumstances of each household and place its card in the appropriate pile. The facilitator did not intervene in this process, and participants held the cards and placed them on the piles themselves. Once all the cards were sorted, the moderator took each pile and read the names back to the informants so they had a chance to review their classification. The moderator then collected the cards from each set, and reclassified them later according to the scheme shown in Table 1.

**Table 1 Setup of database for household food security rating**

| Household name | Group 1 | Group 2 | Group 3 | Group 4 | Group 5 |
|---|---|---|---|---|---|
| Juan Rivera | 1 | 1 | 1 | 2 | 1 |
| Josefina Bobadilla | 2 | 3 | 3 | 3 | 3 |
| Otto Chacón | 1 | 1 | 2 | 2 | 1 |
| etc. | v | w | x | y | z... |

Notes: 1 = Secure; 2 = Intermittently Insecure; 3 = Insecure. Names appearing here are fictitious.

## 3. STATISTICAL ANALYSIS

The aim of the statistical analysis was (1) to assess the level of agreement among different small sets of informants from the same community about the food security status of the families belonging to their beneficiary group, and (2) to examine whether the level of agreement observed in each workshop was associated with other characteristics of the informants (especially their sex) and of the workshop session (especially time elapsed since training of the moderators). The level of agreement between different small sets of informants within the same beneficiary group, averaged over 20 workshop sessions, provided a direct measure of the reliability of the rating method.

Inter-rater agreement on dichotomous outcomes is commonly assessed using a measure referred to as Cohen's Kappa (Fleiss 1981, 217). This statistic ($\kappa$) has the convenient property that $\kappa = +1$ when there is complete agreement between raters,

$0 < \kappa < 1$ when agreement is imperfect but greater than would be expected by chance alone, and $\kappa \leq 0$ when agreement is no more than would be expected by chance. The statistic can be generalized to assess agreement between two raters on an outcome that can take more than two values (Fleiss 1981, 218–219), or agreement between multiple (more than two) raters on a dichotomous outcome (Fleiss 1981, 227). However, no generally accepted method exists for assessing agreement between multiple raters on an outcome that can take more than two values. Therefore, in the current analysis, we assessed separately the agreement between the different sets of raters (informants) on which households in their beneficiary group were "Food Secure," which were "Food Insecure," and which were "Intermittently Insecure." (Note that this was not the case in two of the twenty workshops, where only two small sets of informants were formed and agreement on all three categories was summarized using a single Kappa statistic). As suggested by Landis and Koch (1977), values of $\kappa \geq 0.75$ were considered indicative of excellent agreement beyond that determined by chance; values of $0.4 \leq \kappa < 0.75$ indicated fair to good agreement, and values <0.4 indicated poor agreement.

In order to determine whether the level of agreement between different small sets of informants in the same workshop session was associated with other characteristics of the informants or the session, we used multiple linear regression (McCullagh and Nelder 1989, 48), taking as the dependent variable the Kappa statistics described above. We identified six other variables that we thought likely to affect inter-rater agreement, and included these as independent variables in the regression model. These were

1. **the sex of the informants** (since men might be privy to more personal information about their peers economic conditions than women, or vice versa);
2. **the number of days elapsed since the training session** (since there might be a deterioration or improvement in the quality of the facilitation of the sessions over time);
3. **the HFS category being assessed: "secure", "insecure" or "intermittently insecure"**(since, for example, informants might find it easy to identify which families are food insecure but difficult to determine whether others were secure or intermittently insecure);
4. **the number of sets of informants** (since the statistic used to measure agreement might be sensitive to the number of comparisons made);
5. **the number of subjects being rated** (since smaller groups might have better information flows than bigger groups, facilitating agreement); and
6. **the prevalence (relative frequency in the community) of the category being rated** (since more familiar conditions might be easier for informants to identify than relatively rare ones)

As explained above, each workshop session, except for two, provided three data points: one relating to the raters' agreement on "food secure" households, one relating to their agreement on "food insecure" households, and one relating to their agreement on

"intermittently food insecure" households. Because these three measures were correlated at the level of the workshop session,[4] the analysis was repeated using the Generalized Estimating Equation approach of Liang and Zeger (1986), which models the within-session correlation in the outcome measure as well as the covariate effects, and generally results in more conservative estimates of the regression coefficient standard errors.

The effect of gender on informant behavior may be seen not only in terms of the ability of different sets of informants to agree with one another, but also in terms of bias, a systematic tendency for men and women to classify the same households differently. In order to test this hypothesis, Generalized Estimating Equations models were again used to compare men's and women's informant groups with respect to their propensity to classify individual households as "secure," "insecure," or "intermittently insecure." The model used was equivalent to a standard logistic regression (logit link, Binomial error structure), with an allowance for the fact that different informant sets' assessments of the food security status of a particular household would normally be correlated. This analysis was restricted to the six communities where both men's and women's groups evaluated the HFS status of the same families.

[4] In those communities where the various small sets of informants were in agreement with each other about which households were "secure," they also tended to agree about which households were "insecure" or "intermittently secure." Similarly, poor agreement tended to be consistent across the three categories. This violates the independent errors assumption of Ordinary Least Squares regression, and a correction is needed.

## 4. RESULTS

Findings from the household food security rating exercise are presented in Table 2. A total of 55 intraclass correlation coefficients were derived from the 20 workshop sessions.[5] Thirty-nine coefficients (71 percent) suggested poor agreement between raters, 15 coefficients (27 percent) suggested fair to good agreement, and one coefficient suggested excellent agreement. The median value was 0.27. Figure 1 shows that the agreement varied markedly by class of assignment, with the best agreement obtained for the classification "food secure," and the worst agreement obtained for the classification "intermittently insecure."

The level of agreement also varied in inverse proportion to the number of sets of informants evaluating the same reference subjects (Figure 2). This suggests that the statistic employed did not efficiently normalize for number of sets of informants as expected.

It was further observed that the level of agreement varied in inverse proportion to the number of days elapsed since the exercise was conducted together with the international staff (Figure 3). In a multiple regression analysis (Table 3), the agreement coefficient was estimated to drop by 0.065 (OLS. 95 percent CI: 0.023 to 0.107) with every 10 days elapsed since the training session, and by 0.022 for each additional set of informants (95 percent CI: 0.003 to 0.042).

---

[5]Seven communities had two workshops (one for men; one for women); and six communities had one workshop. This makes for the total of 20 workshops.

**Table 2 Composition of 20 Household Food Security ranking workshops conducted in western Honduras in May/June, 1997. Proportions of households classified as secure/intermittently insecure/insecure, and inter-rater-set agreement (intraclass correlation coefficient)**

| Community | Date | Men/ Women/ Mixed | Number of sets of raters | Number of subjects rated | Percent *secure*[a] | Percent *intermittently insecure*[a] | Percent *insecure*[a] | Agreement *secure* | Agreement *intermittently insecure* | Agreement *insecure* |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | (percent) | (percent) | (percent) | | | |
| El Aguaje | 18 June | Men | 5 | 17 | 12 | 37 | 52 | .43 | .04 | .34 |
| | | Women | 2 | 17 | 20 | 67 | 13 | | .73 | |
| El Campanario | 3 June | Men | 5 | 16 | 23 | 63 | 15 | .78 | .39 | .17 |
| | | Women | 5 | 16 | 9 | 37 | 54 | .61 | .02 | .22 |
| El Ciprés | 13 May | Men | 9 | 26 | 30 | 38 | 32 | .39 | .10 | .35 |
| Chagüite | 14 May | Men | 10 | 19 | 42 | 46 | 12 | .64 | .36 | .11 |
| Chimís Mataras | 13 June | Men | 3 | 14 | 58 | 38 | 5 | .37 | .23 | -.08 |
| | | Women | 2 | 14 | 18 | 61 | 21 | | .27 | |
| La Gocia | 7 May | Mixed | 15 | 22 | 18 | 56 | 27 | .22 | .28 | .63 |
| La Granadilla | 19 June | Men | 8 | 24 | 64 | 33 | 4 | .07 | .02 | -.04 |
| | | Women | 5 | 17 | 22 | 39 | 39 | .16 | .06 | .31 |
| Los Llanos | 15 May | Men | 7 | 32 | 17 | 51 | 31 | .45 | .14 | .30 |
| | | Women | 3 | 32 | 14 | 22 | 64 | .49 | .23 | .53 |
| Mataras | 12 June | Men | 10 | 24 | 24 | 39 | 37 | .00 | .08 | .12 |
| Río Chiquito | 27 May | Men | 4 | 15 | 0 | 49 | 51 | - | .48 | .48 |
| | | Women | 5 | 15 | 15 | 78 | 7 | .08 | .03 | .14 |
| San Marcos | 28 May | Men | 5 | 19 | 29 | 36 | 32 | .30 | .11 | .37 |
| Santa Teresa | 9 May | Men | 11 | 29 | 18 | 69 | 13 | .18 | .21 | .55 |
| Talgua | 2 June | Men | 5 | 30 | 37 | 38 | 25 | .44 | .23 | .41 |
| | | Women | 4 | 31 | 32 | 33 | 35 | .51 | .13 | .40 |

[a] These "percentage" columns represent averages over all sets.

**Figure 1**

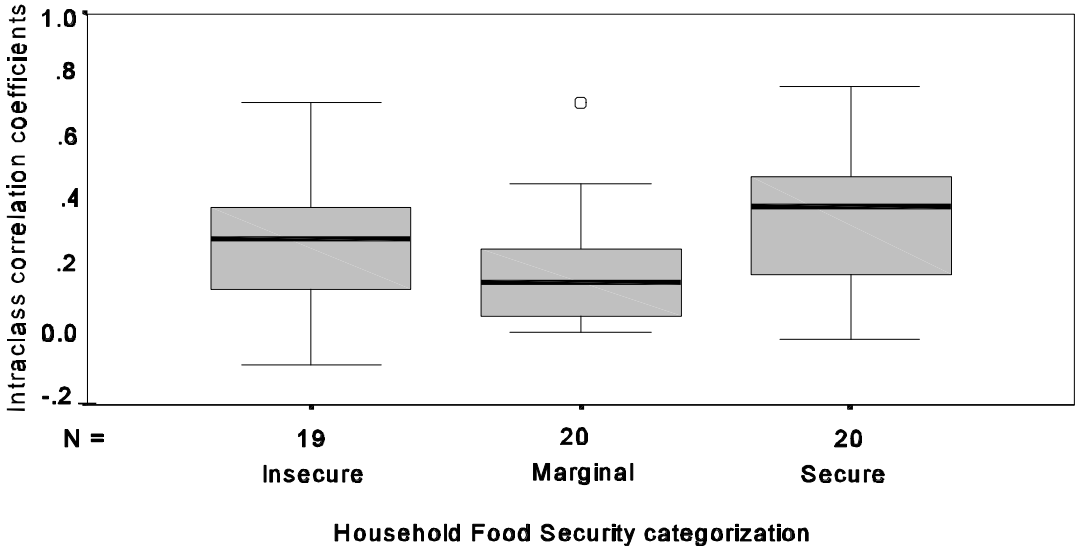


**Figure 2**

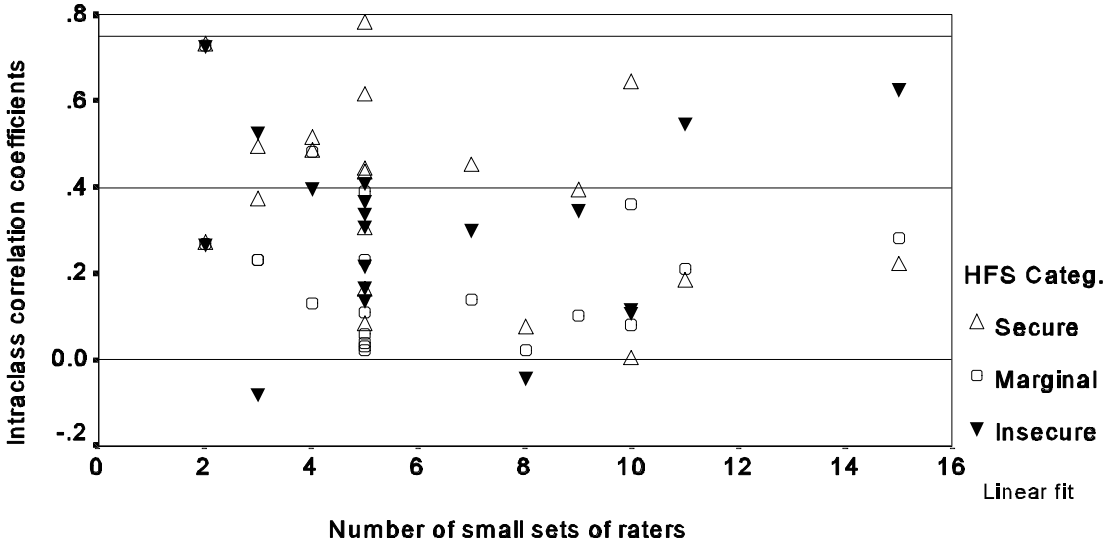

**Figure 3**

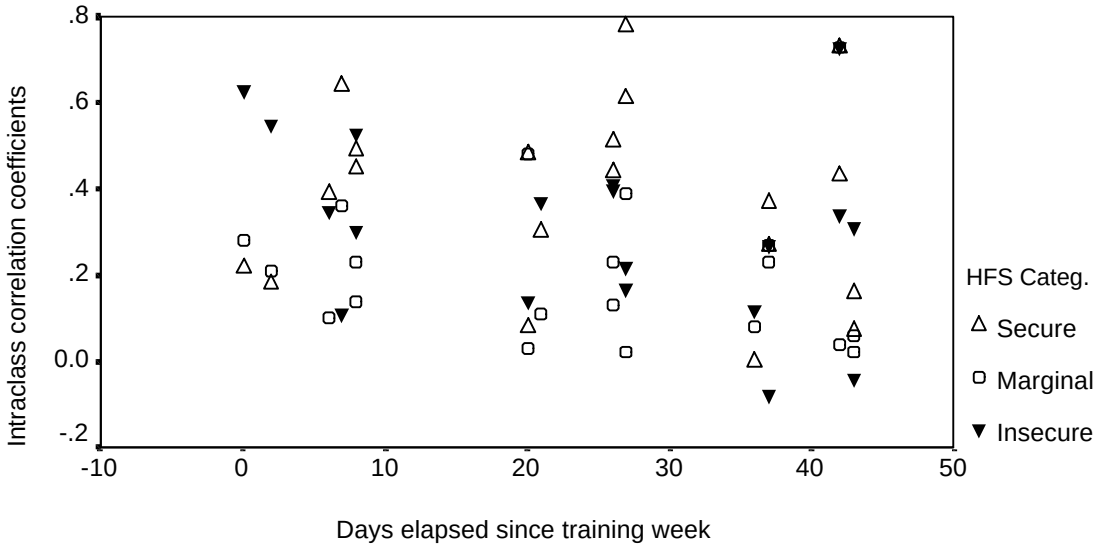
Intraclass correlation coefficients
.8
.6
.4
.2
0.0
-.2
-10
0
10
20
30
40
50
Days elapsed since training week
HFS Categ.
Secure
Marginal
Insecure

**Table 3 The effects of community, session, and household characteristics on the degree of agreement (intraclass correlation coefficient) obtained in the classification of households as food secure, insecure, or intermittently food insecure, western Honduras, 1997 (Ordinary Least Squares and Generalized Estimating Equation regression analysis)**

| | Ordinary Least Squares | | Generalized Estimating Equation | |
|---|---|---|---|---|
| Explanatory variables | Regression coefficient (standard error) | P-value | Regression coefficient (standard error) | P-value |
| Classification as "food insecure" | -0.074 (0.058) | 0.21 | -0.072 (0.051) | 0.16 |
| Classification as "intermittently insecure" | -0.229 (0.064) | 0.001 | -0.224 (0.057) | <0.001 |
| Number of small sets of informants | -0.022 (0.010) | 0.028 | -0.024 (0.014) | 0.095 |
| Total number of subjects rated | -0.002 (0.004) | 0.72 | -0.002 (0.006) | 0.74 |
| Number of days elapsed since training week | -0.007 (0.002) | 0.003 | -0.006 (0.003) | 0.063 |
| Women's group (versus men's/mixed group) | -0.033 (0.059) | 0.57 | -0.003 (0.084) | 0.97 |
| Frequency of HFS class in community | 0.002 (0.002) | 0.11 | 0.002 (0.001) | 0.069 |
| Adjusted $R^2$ = | 0.23 | | | |
| Estimated within-session correlation = | | | 0.36 | |

Note: **Model:** agreement = f (what level of Household Food Insecurity you are agreeing on; how many sets of informants there are; how many subjects you rate; time since training; sex of informants; prevalence of quality that you are rating in your community).

The same analysis indicated that agreement between sets of informants was 0.229 lower for the classification as "intermittently insecure" than for classification as "secure" (95 percent CI: 0.101 to 0.358). There was, however, no significant difference between the classes "secure" and "insecure" (P >.2). The sex of the informants and the number of subjects rated were also unassociated with the level of between-rater agreement (P >0.5). There was some suggestion that the absolute level of food security/insecurity was associated with the degree of agreement obtained (P = 0.11), so that raters would agree more on classifications they were more familiar with in their group. Similar effect estimates were obtained when the data were analyzed using the Generalized Estimating Equations method; however, since this approach results in a smaller effective sample size, the regression coefficient standard errors were generally larger and their corresponding P-values higher.

In the separate analysis of the six communities where men's and women's groups evaluated the HFS status of the same families, men and women were compared with respect to their propensity to classify individuals as "secure," "insecure," or "marginal." Women were 49 percent more likely than men to classify a given family as "insecure" (P <0.001; 95 percent C.I. from +20 percent to +86 percent), and 24 percent less likely to classify a given family as "secure" (P = 0.023; 95 percent C.I. from -40 percent to -4 percent). There was no difference in their propensity to classify a family as "intermittently insecure" (odds ratio = 0.85, P = 0.16). However, examination of Table 2 suggests that this gender effect may not have been constant across communities, since the women's

groups classified more families as food insecure in four villages but less families as food insecure in the remaining two villages. Perhaps what differs between men and women is the *perception* of food security: being contingent on distinctive gender experiences this perception is subjectively defined by each group, thus its consistency within group, but the unpredictable direction of the discrepancy between groups.[6]

## 5. DISCUSSION

The consistency of responses between sets of raters was alarmingly low. The general proposition that group informant ratings can generate reliable information on household wealth should therefore be considered with caution. Many reasons may be invoked to explain why small sets of informants in this study failed to agree on food security rankings. First, the differences between women and men strongly support the hypothesis that social stratification criteria matter. The fact that women and men disagree not only between, but also among, themselves impedes us, however, from seeing a systematic trend; we can only conclude that social factors matter, but in unpredictable ways. Second, the reported loss of homogeneity in group rating as time went by strongly suggests the presence of a study-induced bias, originating in workshop moderator performance. This could indicate a serious gap in training, an issue worth examining in more detail. Third, the different performance of the three HFS categories (the

[6] Note, however, that this explanation still does not explain why, as demonstrated by our other results, ratings are not consistent within groups.

"intermittently insecure" category seemed more difficult to rate than both "secure" and "insecure" categories) suggests a problem with the rating criteria used. The issue may be definitional, or else it may reflect the transitory nature of this category.

Our field experimentation with food security rating exercises yielded several hypotheses that may explain the variation between raters, which are additional to the ones we originally posited. In the remainder of this section, we suggest a number of potential sources of disagreement, and try particularly to distinguish between "real" and "study-induced" biases. These new hypotheses will be addressed in a forthcoming round of field research in different communities of the region.

## 6. HYPOTHESES AS TO SOURCES OF VARIATIONS BETWEEN RATIOS

1. **Different people have different information about the subjects they are rating**. This would correspond to "true" disagreement, and this factor is not something we want to minimize, since it conveys potentially useful information on households being rated. At the same time, there may be aspects that could be controlled to reduce the impact of this factor. Size of the community, for instance, may influence how well people know one another. Also, the reported differences in rating between genders suggest controlling for social factors in group composition. These aspects will be introduced in the coming round of fieldwork.
2. **Different people have different understandings of the criteria used**. The fact that people tended to rate the extreme categories (food secure or food insecure)

more consistently than the intermediary one (intermittently insecure, which seems subjective) suggests that people may find it difficult to fit their fellow villagers in a precoded, binding frame. Particularly, the middle category seems prone to misclassification. This source of variation may be introduced by the study itself, and if so, its design should be revised. It may be more advisable to let people decide themselves about the rating system to adopt and the criteria to use as cutoff points.[7] This alternative will be attempted in the next round of our research to see if a people-defined classification system improves on the consistency of ratings.

3. **Dominance by one or more individuals**: the strength of group ratings is in the range of knowledge that is tapped, and in the overlapping knowledge between informants, combined with successive corrections and approximations through discussion. It is often the case in group settings, however, that one or more individuals dominate the discussion. The more one or a few dominate, the less knowledge is expressed, and the less adjustment there is. This is clearly an issue for facilitators to handle (we return to this point in hypothesis 6). Note, however, that this source of variation is not *necessarily* bad; it could be that the more influential group members are more influential because they have better information. If so, the average correctness of the classification by groups would in fact be improved by this, even if there was more variation between groups because of it. If, on the other

[7]The reason why a precoded format was used in the first round was that food security appeared to us as a dichotomous issue—unlike wealth, which may be viewed as a continuum. For all practical purposes, however, it may be that this distinction between wealth and food security is artificially created by us and that poverty and food insecurity are, in fact, the same from the point of view of villagers.

hand, more influential respondents do not have better information, their dominance will be mainly a source of noise, will not improve the classification, and will preempt adjustment through discussion. There is little that can be done experimentally to control for this, but we may examine whether individual ratings (non-influenced by dominant persons) are, overall, better than group ratings. This will be tests in the next round. We shall do the classification by individuals first, to avoid group ratings influence on individual rankings.

4. **Poor informant selection introduced error in the ratings**: we did not discriminate between informants, using instead all persons present. We agree with Grandin's comment (Grandin 1983) that informant selection is central to the reliability and consistency of ratings. The question remains, however, of how one ever ensures that the right informants are selected. The assumption that one can come up with the "right" set of informants is as subjective as our assumption that knowing one another well ensures consistent ratings. Nevertheless, we will compare group ratings in coming rounds against those offered by a selected number of individuals, as in Grandin's work. We will also compare these two approaches against rankings obtained from formal surveys, to see how these different group compositions perform against the survey approach.[8]

[8] We do not intend to pretend that the survey approach provides an absolute benchmark. It is, however, an alternate form of deriving information, and thus can be used in triangulation with other methods to find the better answer.

5. **Differences at the margin are hard to rate**. The fact that the population was homogeneously "poor" may itself be a problem: small variations in poverty may be weighted differently by distinct informants, thus making the ratings unstable.[9] A test could be done by comparing the consistency of ratings in communities that harbor a wide range of wealth versus those in communities that, like the ones in our group, are already defined along narrow socioeconomic stratification criteria. Unfortunately, the data available from the Honduras study will not allow us to test this hypothesis.

6. **Training of staff was deficient**: as noted, there was a progressive decline in the consistency of ratings as time went by. This source of variation appears to be related to the ways in which the workshops are conducted, rather than to true differences of opinion among the informants. Facilitator bias (introduced by preconceived opinions, by increasing expediency in handling the workshops, by a poor presentation of rating criteria, or by simple boredom) may exert powerful effects on the attitude of informants and the seriousness with which they consider the exercise. The implication is that the greatest care should go into preparation and training of facilitators.

[9] A similar problem was noted by Gujit who mentioned that "minute differences may make the distinction irrelevant" (Gujit 1992b).

## 7. CONCLUSION

We believe that informant rating methods can provide great services to field researchers. This paper suggests, however, that special precautions must be taken when using this type of technique, as the reliability of ratings seems to vary according to many factors. Some of the variation undoubtedly refers to genuine informant disagreement; but, also, the study design may be itself the source of discrepancies. Aspects of the method most likely to introduce bias include staff training, choice of ranking criteria, informant selection, and definition of the target population. More research is needed to clarify these issues. In the meantime, it is suggested that great care be exercised when using this type of approach.

## REFERENCES

Adams, A. M., T. G. Evans, R. Mohammed, and J. Farnsworth. 1997. Socioeconomic stratification by wealth ranking: Is it valid? *World Development* 25 (7): 1165–1172.

Briefel, R. R., and C. E. Woteki. 1992. Development of food sufficiency questions for the Third National Health and Nutrition Examination Survey. *Journal of Nutrition Education* 24 (Supplement 1): 24S–28S.

Carmines, E. G., and R. A. Zeller. 1983. *Reliability and validity assessment. Series Quantitative applications in the social sciences*. Sage University Paper No. 17. 6th edition. London: Sage Publications.

Chadwick, M. T., J. A. Seeley, and G. R. Sherchan. 1995. Preliminary comments on a wealth ranking exercise performed with forest user groups from three geographical areas of Nepal. Nepal-U.K. Forestry Research Project, Kathmandu.

Chambers, R. 1997. *Whose reality counts?: Putting the first last*. London: Intermediate Technology.

FAO (Food and Agriculture Organization of the United Nations). 1992. *Sociological analysis in agricultural investment project design*. Rome.

Fleiss, J. L. 1981. *Statistical methods for rates and proportions*. New York: John Wiley and Sons.

Grandin, B. 1983. The importance of wealth in pastoral production: A rapid method of wealth ranking. In *Pastoral systems research*. Addis Ababa: International Livestock Centre for Africa (ILCA).

Grandin, B. 1988. Wealth ranking in smallholder communities: A field manual. London: IT Publications.

Grandin, B. 1994. Wealth ranking. In *Tools for the field: Methodologies handbook for gender analysis in agriculture*, ed. H. S. Feldstein and J. Jiggins. West Hartford, Conn., U.S.A.: Kumarian Press.

Gujit, I. 1992. The elusive poor. A wealth of ways to find them. Report on the Institute of Development Studies/International Institute for Environment and Development (IDS/IIED) Seminar on wealth and well-being ranking. *RRA Notes* 15. London: International Institute for Environment and Development.

Gujit, I. 1992. A user's note on wealth ranking by cards and list of sources on wealth ranking. *RRA Notes* 15. London: International Institute for Environment and Development (IIED).

IIED (International Institute for Environment and Development). 1992. Special issue on wealth ranking. *RRA Notes* 15. London.

Isham, J., D. Narayan-Parker, and L. Pritchett. 1994. *Does participation improve performance?: Empirical evidence from project data*. Washington, D.C.: World Bank.

Jacobs, S. 1996. Social assessment and participation. Methods and tools. Draft. World Bank, Washington, D.C.

Landis, J. R., and G. G. Koch. 1974. The measurement of observer agreement for categorical data. *Biometrics* 33: 159–174.

Liang K.-Y., and S. L. Zeger. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73: 13–22.

McCullagh, P., and J. A. Nelder. 1989. *Generalized linear models*. 2nd edition. London: Chapman & Hall.

Mearns, R., and G. Shombodon. 1992. Direct and indirect uses of wealth ranking in Mongolia. *RRA Notes* 15. London: International Institute for Environment and Development (IIED).

Mukherjee, N. 1992. Villagers' perceptions of rural poverty through the mapping methods of PRA. *RRA Notes* 15. London: International Institute for Environment and Development (IIED).

Pretty, J., S. Subramanian, N. Kempu Chetty, D. Ananthakrishnan, C. Jayanthi, S. Muralikrishinasamy, and K. Renganayaki. 1992. Finding the poorest in a Tamil Nadu village. A sequence of mapping and wealth ranking. *RRA Notes* 15. London: International Institute for Environment and Development (IIED).

Riely, F., N. Mock, B. Cogill, L. Bailey, and E. Kenefick. 1995. Food security indicators and framework for use in the monitoring and evaluation of food aid programs. IMPACT Project. U.S. Agency for International Development, Washington, D.C.

Sarch, M. T. 1992. Wealth ranking in The Gambia. Which households participated in the FITT Programme? *RRA Notes* 15. London: International Institute for Environment and Development (IIED).

Schaefer, S. 1992. The “Beans-game”—Experiences with a variation of wealth ranking in the Kivu region, Eastern Zaire. *RRA Notes* 15. London: International Institute for Environment and Development (IIED).

Wellbourn, A. 1992. A note on the use of disease problem ranking with relation to socioeconomic well-being: An example from Sierra Leone. *RRA Notes* 16. London: International Institute for Environment and Development (IIED).